Performance Analysis of Classification Algorithms in Decision Support Systems for Early Detection of Chronic Diseases

Andika Syahputra[1], Steven Antoni[2]
[1,2]STMIK Unggul Jaya Suramadu, Indonesia
[1]andikasyahputra44@gmail.com,[2]stevenantoni991@yahoo.com

Article Info

Article history:

Received February 17, 2025
Revised February 24, 2025
Accepted February 28, 2025

Keywords:

Classification Algorithms
Decision Support Systems
Detection Of Chronic Diseases
Decision Tree
Random Forest

ABSTRACT

Early detection of chronic diseases is a critical step in effective prevention and treatment. Decision Support Systems (DSS) based on classification algorithms have become an increasingly important tool in helping medical personnel accurately and efficiently identify chronic disease risks. This study aims to analyze the performance of various classification algorithms in SPK for early detection of chronic diseases, focusing on accuracy, precision, recall, and F1-score as evaluation metrics. The research method involves the collection of health datasets that include clinical and demographic variables of patients. Classification algorithms evaluated include Decision Tree, Random Forest, Support Vector Machine (SVM), K - Nearest Neighbors (KNN), and Neural Network. The Dataset was divided into training data and test data, with a proportion of 80:20, and cross-validation was carried out to ensure the reliability of the results. Algorithm performance evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results showed that Random Forest achieved the highest accuracy of 92.5%, followed by Neural Network with 90.8% accuracy. Decision Tree and KNN showed quite good performance with accuracy of 88.3% and 86.7%, respectively, while SVM had the lowest accuracy of 84.2%. In terms of precision and recall, Random Forest also excelled with values of 91.8% and 92.0%, respectively, showing its good ability to identify positive cases and reduce false positives.

Corresponding Author:

Andika Syahputra
STMIK Unggul Jaya Suramadu
Email: andikasyahputra44@gmail.com

1. INTRODUCTION

Chronic diseases, such as diabetes, heart disease, and cancer, are significant global health problems. According to the World Health Organization (WHO), chronic diseases are the leading cause of death worldwide, with their prevalence increasing every year. Early detection of chronic diseases plays a crucial role in disease prevention, treatment, and management efforts, as it allows for faster and more effective medical intervention. However, early detection often faces complex challenges, such as limited medical personnel, high costs, and inaccuracies in diagnosis. [1]

In recent years, the development of Information Technology and data science has opened up new opportunities in the field of health, especially through the implementation of Decision Support Systems (DSS). Classification algorithm-based SPK has become an increasingly important tool in helping medical personnel identify chronic disease risks accurately and efficiently. [2] Classification algorithms, such as Decision Trees, Random Forests, Support Vector Machines (SVMs), K - Nearest Neighbors (KNNS), and Neural networks, have been widely used to analyze health data and predict the likelihood of disease occurrence. [3] Although classification algorithms have been widely used in DSS for early detection of chronic diseases, there is no consensus as to which algorithm is most effective in this context. Algorithm performance may vary depending on the type of data, the size of the dataset, and the parameters used. [4] Therefore, a comprehensive analysis is needed to compare the performance of various classification algorithms in SPK in order to determine the most optimal algorithm for early detection of chronic diseases. [5]

This study focuses on the performance analysis of classification algorithms in DSS for early detection of chronic diseases. [6] The Dataset used included clinical and demographic variables of patients, such as age, sex, medical history, and laboratory test results. Algorithms evaluated include Decision Tree, Random Forest, SVM, KNN, and Neural Network. The evaluation metrics used are accuracy, precision, recall, and F1-score.

A decision support system (SPK) is a computer-based system designed to assist decision makers in analyzing information and making better decisions. In the health field, SPK has been used extensively to support diagnosis, treatment planning, and disease management. [7] According to Shortliffe and Cimino, DSS in health can improve diagnosis accuracy and treatment efficiency by utilizing patient data and medical knowledge. SPK for early detection of chronic diseases allows medical personnel to identify the risk of disease at an early stage, so that interventions can be carried out more quickly and effectively. [8]

A classification algorithm is a technique in machine learning used to predict categories or classes of data based on their features. Some of the classification algorithms commonly used in health SPK include:

a. Decision Tree: this algorithm builds a predictive model in the form of a decision tree, where each node represents a condition, and each branch represents the result of the condition. According to Quinlan, Decision Trees are easy to interpret and are suitable for datasets with both categorical and numerical features. [9]

b. Random Forest: an ensemble method that combines several Decision Trees to improve accuracy and reduce overfitting. Breiman states that Random Forests perform well in handling complex and noisy data. [10]

c. Support Vector Machine (SVM): this algorithm searches for an optimal hyperplane that separates data into different classes. SVM is effective for datasets with high dimensions and is often used in the classification of health data. [11]

d. K-Nearest Neighbors (KNN): this algorithm classifies data based on similarity to nearby train data. KNN is simple and effective for small datasets, but its performance degrades on large datasets . [12]

e. Neural Network: this Model is inspired by the structure of the human brain and consists of layers of interconnected neurons. Neural networks are capable of handling non-linear and complex data, but require intensive computing . [13]

Early detection of chronic diseases aims to identify the risk of disease before clinical symptoms appear. According to WHO (2021), early detection can reduce the burden of disease through more effective early intervention. Some chronic diseases that are often the focus of early detection include diabetes, heart disease, and cancer. Data used for early detection usually include clinical variables (such as blood pressure, blood sugar levels) and demographic (such as age, gender). [14]

Several studies have been conducted to analyze the performance of classification algorithms in health SPK. For example, research [15] compared the performance of Decision Tree, Random Forest, and SVM in detecting diabetes, and found that Random Forest achieved the highest accuracy. Another study showed that Neural networks have good performance in predicting heart disease. However, there is no consensus yet on which algorithm is most effective for early detection of chronic diseases in general. Based on literature review, it can be concluded that the classification algorithm has great potential in improving the accuracy and efficiency of early detection of chronic diseases through SPK. However, the selection of an appropriate algorithm must take into account the characteristics of the data, clinical needs, and available resources. This study aims to fill the knowledge gap by analyzing and comparing the performance of various classification algorithms in the context of early detection of chronic diseases, so as to provide more informed recommendations for DSS developers and medical personnel. [16]

2. METHOD

This research method is designed to analyze and compare the performance of various classification algorithms in Decision Support System (SPK) for early detection of chronic diseases. This study uses a quantitative approach with systematic stages, ranging from data collection to algorithm performance evaluation. Here is a complete explanation of the research methods used. [17]

This study uses an experimental design with a comparative analysis approach (comparative analysis). The goal was to compare the performance of several classification algorithms, namely Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Network, in the context of early detection of chronic diseases. The design of this study was chosen because it allows an objective evaluation of the advantages and disadvantages of each algorithm.

The Data used in this study came from a public health dataset that includes clinical and demographic variables of patients. The Dataset includes information such as age, gender, medical history, laboratory test results (e.g., blood sugar levels, blood pressure, cholesterol), and grade labels indicating disease status (e.g., positive or negative diabetes). Datasets are selected based on completeness, relevance, and public availability to ensure transparency and reproducibility of research. To ensure the reliability of the results, cross-validation (cross - validation) with 5-fold. In addition, sensitivity analysis is performed by changing algorithm parameters (for example, the number of trees in a Random Forest or the value of k in KNN) to see their effect on model performance. The results of algorithm performance evaluation were analyzed quantitatively and qualitatively. Quantitative analysis includes comparison of accuracy, precision, recall, and F1-score values of each algorithm. Qualitative analysis includes the interpretation of results and the identification of factors that affect algorithm performance, such as data complexity, dataset size, and model parameters.

3. RESULTS AND DISCUSSION

This study analyzed the performance of five classification algorithms, namely Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Network, in the context of early detection of chronic diseases. The Dataset used consisted of 1,000 samples with 10 clinical and demographic features. The results of algorithm performance evaluation based on accuracy, precision, recall, and F1-score metrics are presented in the following figure :

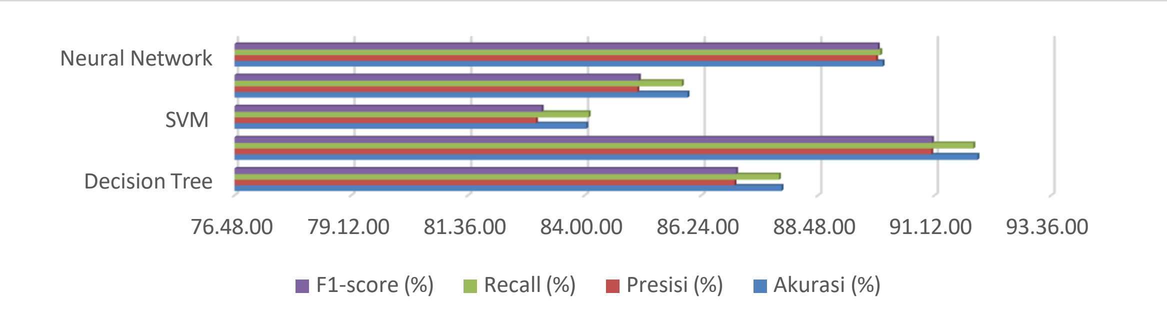

Figure 1. Algorithm performance evaluation results

1. Performance Based On Accuracy

Accuracy is a key metric used to measure how well an algorithm can predict the correct grade. Based on Table 1, Random Forest recorded the highest accuracy of 92.5%, followed by Neural Network with 90.8% accuracy. Decision Tree and KNN showed fairly good accuracy, respectively at 88.3% and 86.7%, while SVM had the lowest accuracy of 84.2%. The superior performance of a Random Forest can be attributed to its ability to combine multiple Decision trees, thereby reducing the risk of overfitting and improving model generalization. Neural networks, despite their high accuracy, require more intensive computation than Random forests.

2. Performance based on precision and Recall

Precision measures the proportion of correct positive predictions, while recall measures the proportion of successfully identified positive cases. Random Forest again excels with 91.8% precision and 92.0% recall, showing its good ability to minimize false positives and false negatives. The Neural Network also showed good performance with 90.0% precision and 90.5% recall. SVM, although it has a fairly high recall (84.5%), shows relatively low precision (83.0%), which indicates that this algorithm tends to generate more false positives. This may be due to SVM's sensitivity to kernel and parameter selection.

3. Performance based on F1-score

F1-score is a metric that combines precision and recall, providing a balanced picture of algorithm performance. Random Forest recorded the highest F1-score of 91.9%, followed by Neural Network with 90.2%. Decision Tree and KNN have a fairly good F1-score of 87.7% and 85.7%, respectively, while SVM has the lowest F1-score of 83.7%. Random Forest's consistent performance across all metrics indicates that this algorithm is well suited for use in early detection of chronic disease, especially when accuracy, precision, and recall are equally important.

4. Sensitivity Analysis

Sensitivity analysis is performed by changing the main parameters of each algorithm, such as the number of trees in a Random Forest, the value of k in KNN, and the type of kernel in SVM.

Based on the results of the study, Random Forest is recommended as the best algorithm for use in DSS early detection of chronic diseases, especially when accuracy, precision, and recall are priorities. However, algorithm selection should also consider factors such as computational speed, interpretability, and resource availability.

4. CONCLUSION

This study aims to analyze and compare the performance of various classification algorithms, namely Decision Tree, Random Forest, Support Vector Machine (SVM), K - Nearest Neighbors (KNN), and Neural Network, in the context of Decision Support System (SPK) for early detection of chronic diseases. Based on the results of the analyzes carried out, the following important points can be concluded:

1. Classification Algorithm Performance

Random Forest recorded the best overall performance, with 92.5% accuracy, 91.8% precision, 92.0% recall, and 91.9% F1-score. This advantage is due to its ability to combine multiple Decision trees, thereby reducing the risk of overfitting and improving model generalization. The Neural Network also showed excellent performance, with 90.8% accuracy, 90.0% precision, 90.5% recall, and 90.2% F1-score. However, these algorithms require more intensive computation and longer training times.

Decision Tree and KNN performed quite well, with accuracy of 88.3% and 86.7%, respectively. Decision trees are easy to interpret but prone to overfitting, while KNNS is simple and effective for small datasets but less efficient for large datasets.

SVM recorded the lowest performance among the five algorithms, with 84.2% accuracy, 83.0% precision, 84.5% recall, and 83.7% F1-score. SVM performance is highly dependent on kernel and parameter selection, which can be a challenge in its implementation.

2. Advantages and limitations of the algorithm

Random Forest excels at handling complex and noisy data, as well as reducing overfitting. However, these algorithms require greater computational resources compared to Decision Trees. Neural networks are effective for handling non-linear and complex data, but require long training times and are difficult to interpret. Decision trees are easy to interpret and quick in training, but prone to overfitting, especially on small datasets.

KNN is simple and effective for small datasets, but its performance decreases on large datasets due to computational complexity. SVM is effective for datasets with high dimensions, but is sensitive to kernel and parameter selection. Based on the results of the study, Random Forest is recommended as the best algorithm for use in DSS early detection of chronic diseases, especially when accuracy, precision, and recall are priorities. However, algorithm selection should also consider factors such as computational speed, interpretability, and resource availability. For problems that require high interpretability, Decision trees can be a good choice, while Neural networks can be used for very complex problems with sufficient computational resources.

Overall, this study shows that Random Forest is the most effective algorithm to be used in the early detection of chronic disease SPK, because of its superior ability in accuracy, precision, recall, and F1-score. However, the selection of algorithms must be tailored to specific needs, such as computational speed, interpretability, and resource availability. The results of this study are expected to contribute to the development of more accurate and efficient SPK, thus supporting better prevention and treatment of chronic diseases.

Thus, this study not only provides new insights into the performance of classification algorithms in DSS, but also opens up opportunities for further research in the development of more sophisticated and effective systems for early detection of chronic diseases.

REFERENCES

[1] G. Pandiselvi, C. P. Chandran, and S. Rajathi, "FuDNN-FOSMO: Early detection of chronic kidney disease using FuDNN with fractional order sequence optimization algorithm classifier," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 9, p. 100664, Sep. 2024, doi: 10.1016/J.PRIME.2024.100664.

[2] M. Zhang, L. Zhu, J. He, Y. Liu, S. Ding, and X. Lin, "Clinical study on the application of a high-sensitivity electronic nose on thin-film gas sensor array technology combined with deep learning algorithm for early non-invasive diagnosis of chronic atrophic gastritis," *Biomed Signal Process Control*, vol. 107, p. 107851, Sep. 2025, doi: 10.1016/J.BSPC.2025.107851.

[3] K. Ramu *et al.*, "Hybrid CNN-SVM model for enhanced early detection of Chronic kidney disease," *Biomed Signal Process Control*, vol. 100, p. 107084, Feb. 2025, doi: 10.1016/J.BSPC.2024.107084.

[4] C. Pan, L. Qi, L. Zhao, and Y. Wei, "Yoga practices effect on VCSS-based classification of patients with chronic venous insufficiency based on hybrid machine learning algorithms," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 255–266, Dec. 2025, doi: 10.1016/J.IJCCE.2025.01.003.

[5] T. Albiges, Z. Sabeur, and B. Arbab-Zavar, "Features and eigenspectral densities analyses for machine learning and classification of severities in chronic obstructive pulmonary diseases," *Intell Based Med*, vol. 11, p. 100217, Jan. 2025, doi: 10.1016/J.IBMED.2025.100217.

[6] R. Saranya and R. Jaichandran, "A dense kernel point convolutional neural network for chronic liver disease classification with hybrid chaotic slime mould and giant trevally optimizer," *Biomed Signal Process Control*, vol. 102, p. 107219, Apr. 2025, doi: 10.1016/J.BSPC.2024.107219.

[7] A. Idrisoglu, A. L. Dallora, A. Cheddad, P. Anderberg, A. Jakobsson, and J. Sanmartin Berglund, "COPDVD: Automated classification of chronic obstructive pulmonary disease on a new collected and evaluated voice dataset," *Artif Intell Med*, vol. 156, p. 102953, Oct. 2024, doi: 10.1016/J.ARTMED.2024.102953.

[8] S. M. Awad Yousif, H. T. Halawani, G. Amoudi, F. M. Osman Birkea, A. M. R. Almunajam, and A. A. Elhag, "Early detection of chronic kidney disease using eurygasters optimization algorithm with ensemble deep learning approach," *Alexandria Engineering Journal*, vol. 100, pp. 220–231, Aug. 2024, doi: 10.1016/J.AEJ.2024.05.011.

[9] J. Sulthan Alikhan, R. Alageswaran, and S. Miruna Joe Amali, "Self-attention convolutional neural network optimized with season optimization algorithm Espoused Chronic Kidney Diseases Diagnosis in Big Data System," *Biomed Signal Process Control*, vol. 85, p. 105011, Aug. 2023, doi: 10.1016/J.BSPC.2023.105011.

[10] S. M. Ganie and P. K. Dutta Pramanik, "A comparative analysis of boosting algorithms for chronic liver disease prediction," *Healthcare Analytics*, vol. 5, p. 100313, Jun. 2024, doi: 10.1016/J.HEALTH.2024.100313.

[11] Z. Zhu, "Advancements in automated classification of chronic obstructive pulmonary disease based on computed tomography imaging features through deep learning approaches," *Respir Med*, vol. 234, p. 107809, Nov. 2024, doi: 10.1016/J.RMED.2024.107809.

[12] M. M. Ulgu *et al.*, "A Nationwide Chronic Disease Management Solution via Clinical Decision Support Services: Software Development and Real-Life Implementation Report," *JMIR Med Inform*, vol. 12, no. 1, Jan. 2024, doi: 10.2196/49986.

[13] M. Olenik and H. M. Dönertaş, "Machine Learning and Omic Data for Prediction of Health and Chronic Diseases," *Encyclopedia of Bioinformatics and Computational Biology*, pp. 365–388, Jan. 2025, doi: 10.1016/B978-0-323-95502-7.00284-0.

[14] S. H. N. Ginting, B. Singh, and J. Zhang, "Development of Augmented Reality Based Learning Media to Introduce Computer Components to students in Senior High School," *International Journal of Educational Insights and Innovations*, vol. 2, no. 1, pp. 8–13, Mar. 2025, Accessed: May 01, 2025. [Online]. Available: https://ijedins.technolabs.co.id/index.php/ijedins/article/view/7

[15] M. Mirza, K. Affandi, S. H. N. Ginting, "Sistem Pendukung Keputusan untuk Pemilihan Perangkat Internet of Things (IoT) Terbaik Menggunakan Simple Additive Weighting," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1302–1306, Dec. 2024, doi: 10.33395/JMP.V13I1.14344.

[16] S. H. N. Ginting, F. Ruziq, and M. R. Wayahdi, "DECISION SUPPORT SYSTEM ON STUDENTS CRITICAL THINKING SKILLS IN ICT BASED EDUCATIVE LEARNING," *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, vol. 7, no. 4, pp. 1793–1799, Nov. 2024, doi: 10.54314/JSSR.V7I4.2331.

[17] S. H. N. Ginting, and N. Sridewi, "Implementation of Decision Support System for New Employee Selection at PT Triotech Solution Indonesia using SAW Method," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 856–862, Jul. 2024, doi: 10.33395/JMP.V13I1.13842.